N° TSE-494

March 2014

« Spatial dependence in (origin-destination) air passenger flows »

Romain Doucet, Paula Margaretic,

Christine Thomas-Agnan and Quentin Villotta

Spatial dependence in (origin-destination) air passenger flows*

Romain Doucet †Paula Margaretic ‡ Christine Thomas- Agnan§ Quentin Villotta ¶

March 2014

Abstract

We explore the estimation of origin-destination (OD), city-pair, air passengers, in order to explicitly take into account spatial autocorrelation. To our knowledge, we are the first to test the presence of spatial autocorrelation and apply spatial econometric OD flow models to air transport. Drawing on a world sample of 279 cities, over $2010-2012$, we find significant evidence of spatial autocorrelation in air passenger flows. Thus, contrary to common practice, we need to incorporate the spatial structure present in the data, when estimating OD air passengers. Importantly, failure to do it, may lead to inefficient estimated coefficients and prediction bias.

Keywords: Spatial autocorrelation, spatial econometric origin-destination flow model, air passenger flows.

*We would like to gratefully acknowledge the head of Market Research and Forecasts department at Airbus, David Prevor, for his constant support during this project; Fabrice Valentin and the Market Research and Forecasts team, Lionel Cousseins, Ismail Ibrahim, Guillaume Legeay, Francois Xavier Le-Goff, Josh Myers, for very interesting discussions. Marco Batarce, Thibault Laurent and Jean Philippe Lesne provided helpful suggestions. All remaining errors are naturally ours.

†Airbus SAS.

‡Airbus SAS. Contact: paula.margaretic@airbus.com.

§Toulouse School of Economics. Contact: christine.thomas@tse-fr.eu.

¶Airbus SAS.

1 Introduction

This paper investigates the estimation of origin-destination (OD) air passenger volume. Our interest is to estimate air passenger traffic from one city to another. Among the factors that make a city attractive for passengers, the literature has mainly focused on the size of the city population and its socioeconomic development, as measured, for example, by income per capita. However, much less attention has been given to the spatial dependence among these factors.

Spatial dependence means the co-variation of factors within a geographic space. In our context, this implies that the characteristics at proximal cities may impact air passenger flows, between two cities.[1] Because spatial dependence violates the typical independence assumption made in regression analysis, our aim is to study whether and how spatial dependence plays a role, when estimating OD air passengers. Importantly, failure to properly account for spatial dependence, when it exists, may lead to inefficient estimated coefficients and prediction bias, among others.

Our paper has three main motivations. The first one is empirical: Spatial interaction models focus on OD flow data. Among them, gravity models have been extensively used, with numerous applications in trade, migration and air transportation.[2] The main particularity of gravity models is that they rely on a function of the distance between origin and destination (together with characteristics of both origins and destinations), assuming that distance can effectively eliminate the spatial dependence potentially present in OD flow data.

However, numerous investigations have challenged this assumption, both theoretically and empirically.[3] A prolific strand of literature has emerged, proposing alternative ways to extend spatial interaction models to account for spatial dependence. Among others, Lesage and Pace (2008) propose to incorporate spatial autoregressive dependence (spatial lag), while Dubin (2003) works with a spatially auto correlated error term (spatial error).[4] Our motivation is to contribute to this debate and assess whether these two forms of spatial structure play any role, when estimating OD air passenger flows.

Second, from an applied point of view, being able to estimate the number of air passengers between two cities at a given point in time is of major importance both for aircraft manufacturers and airlines. Aircraft manufacturers, such as Airbus, rely on this type of modeling to assess the future demand for civil passenger and freighter aircraft, which in turn, steer them towards innovation. Airlines also need these models to decide whether to open new routes, offer more

[1] See Lesage and Pace (2008) for a discussion.

[2] Bhadra and Kee (2008), Doganis (2004), Jorge-Calderon (1997) and Russon and Riley (1993) are examples of the application of gravity models to air transport. See Grosche *et al.* (2007) for a literature review.

[3] Curry (1972) has been the first to argue that spatial autocorrelation effects are confounded with distance decay effects during the estimation of gravity model parameters. In turn, using journey-to-work data, Griffith and Jones in 1980 show that spatial autocorrelation matters. Tiefelsdorf in 2003 arrives to the same conclusion, using migration flow data.

[4] See also Dubin (2004) and Lesage and Pace (2004 and 2010).

frequencies and/or increase aircraft capacity.

Finally, from a policy standpoint, better predicting OD air passengers can also be useful for airport planners, government and non-government agencies and air transport and economic policy-makers world-wide. As an illustration, since the 1979 Airline Act Deregulation in the US, there has been a global trend towards liberalization of air travel in Europe, Asia and Latin America. There is now a strong need to evaluate the impact of these regional measures on air traffic. Properly accounting for spatial interactions can help us better evaluate the effect of these policies.

Drawing on a sample of 279 cities around the world over the period $2010 - 2012$, we first apply the traditional gravity model to estimate air passenger flows. Second, we test the presence of spatial autocorrelation. Third, inspired by Dubin (2003) and Lesage and Pace (2008), we introduce two spatial connectivity matrices, for origin and destination spatial dependence, and modify the gravity model to account for spatial dependence, both in air passenger flows and the disturbances.

To our knowledge, we are the first to test the presence of spatial autocorrelation and apply spatial econometric models that account for spatial dependence to air transport.[5] Another virtue of our application is that the dataset is global, that is, the 279 cities belong to the five continents.

We estimate six spatial models, which allow for spatial autoregressive dependence or spatially auto correlated error term. We then compare each of these six spatial models with the gravity model, which assumes no spatial dependence. Based on Akaike informational criteria and likelihood ratio tests, we conclude that any spatial model is better than the least-square one.

This result has two key implications. First, we need to incorporate the spatial patterns of the geographical phenomena, when estimating OD air passengers. Second, despite the common practice, least-square estimates and inferences that ignore this spatial dependence in air transport seem not to be justified.

The paper closest to ours is Lesage and Pace (2008). They propose a way to incorporate spatial autoregressive dependence to the traditional gravity model. We extend their model, by allowing for a spatially autocorrelated error term. We apply their technical results to air transport and conclude that spatial dependence matters when estimating air passenger flows.

The paper proceeds as follows. Section two introduces the traditional gravity model we consider here and the modification to account for spatial dependence, both in air passenger flows and the disturbances. Section three presents the data set. Section four shows the estimate results, first assuming independent observations and then allowing for spatial dependence. Section five discusses one application of this type of modeling to air transport. Concluding remarks are in section six. Additional estimate results and robustness checks are relegated to the Appendix.

[5] By calibrating a gravity model for 100 American cities in 1970, Fotheringham (1981) shows evidence of the relationship between distance decay parameters and the size and configuration of origins and destinations. Boros *et. al* in 1993 test the presence of spatial autocorrelation, using data on daily flights for nine main airlines operating in US domestic market in 1992.

2 Spatial interaction model for (OD, city pair) air passenger flows

Section 2 starts by introducing the notation needed to model (OD, city-pair) air passenger flows. Second, it presents one type of spatial interaction model, the square gravity model, assuming independent observations. Third, following Dubin (2003) and Lesage and Pace (2008), it introduces two spatial connectivity matrices for origin and destination spatial dependence and modifies the spatial interaction model to account for spatial dependence, both in air passenger flows and the disturbances.

2.1 Air passenger flows

At any time period t, let $\mathbf{Y}_t$ be an $n \times n$ matrix of air passenger flows, where the n columns represent cities of origin (o) 1 to n and the n rows correspond to destination cities (d) 1 to n :

$$\mathbf{Y}_t = \begin{pmatrix} o_1 \rightarrow d_1 & o_2 \rightarrow d_1 & ... & o_n \rightarrow d_1 \\ o_1 \rightarrow d_2 & o_2 \rightarrow d_2 & ... & ... \\ & & & o_n \rightarrow d_{n-1} \\ & & & o_n \rightarrow d_n \end{pmatrix} \tag{1}$$

As in Lesage and Pace (2008), we can create an $N \times 1$ vector of air passenger flows, with $N = n^2$, from the flow matrix (1) in two ways: an origin-centric ordering or a destination-centric ordering. Denote $\mathbf{y}_t$ the $N \times 1$ air passenger flow vector. An origin-centric ordering requires $\mathbf{y}_t^o = vec(\mathbf{Y}_t)$, whereas a destination-centric ordering needs $\mathbf{y}_t^d = vec(\mathbf{Y}_t')$.

Without loss of generality, hereafter, we focus on the origin-centric ordering, hence $\mathbf{y}_t = \mathbf{y}_t^o$, with the first n rows of y_t corresponding to air passengers from origin 1 to all the n destination cities at period t, while the last n rows of $\mathbf{y}_t$ referring to air passengers from city of origin n to all the n destination cities, also at t. For brevity, hereafter, we omit the subindex t.

2.2 Gravity model with independent observations

The square ($n^2 = N$) gravity model we study here relates average air passenger flows to the origin and destination city characteristics. Also, it models interdependence among observations using distance.

Define $\mathbf{X}$ as the $n \times k$ matrix of explanatory variables, containing k characteristics of the n cities. Given the $N \times 1$ vector of air passenger flows, $\mathbf{y}$, we need to repeat $\mathbf{X}$ n times to create an $N \times k$ matrix, that we label $\mathbf{X}_d$, which contains the characteristics of the destination cities. Hence, $\mathbf{X}_d = \mathbf{i}_n \otimes \mathbf{X}$, with $\mathbf{i}_n$ an $n \times 1$ unit vector and $\otimes$ the Kronecker product. Similarly, we define the $N \times k$ matrix of origin characteristics as $\mathbf{X}_o = \mathbf{X} \otimes \mathbf{i}_n$.[6]

[6] $\mathbf{X}_o$ repeats the characteristics of the origin city 1, n times to form the first n rows of $\mathbf{X}_o$; the characteristics of

Let $\mathbf{G}$ be an $n \times n$ matrix of distances between origins and destinations and $\mathbf{g} \equiv vec(\mathbf{G})$ is a $N \times 1$ vector of these distances from each city of origin to each destination city.

The least square regression of the N gravity model becomes,

$$\mathbf{y} = \alpha i_N + \beta_d \mathbf{X}_d + \beta_o \mathbf{X}_o + \gamma \mathbf{g} + u, \tag{2}$$

with αi_N an $N \times 1$ constant parameter vector, β_d and β_o the $k \times 1$ parameter vectors and γ the scalar distance parameter. Finally, we assume for the moment the $N \times 1$ error vector as $u \sim N(0, \sigma^2 \mathbf{I}_N)$.

2.3 Spatial dependence

The previously defined gravity model assumes independence among observations. However, this assumption may be inadequate in many applications (see Griffith (2007) for a discussion). Moreover, the failure to consider spatial dependence may lead to inefficient estimated coefficients and prediction bias, among others.

In order to account for spatial dependence, we start by introducing the neighbourhood weight matrix, $\mathbf{W}$.[7] The m-nearest neighbour weight matrix $\mathbf{W}$ represents a $n \times n$ non-negative, sparse matrix, with element $w_{ij} > 0$ if city i is one of the m-nearest neighbours to city j and $\sum_j w_{ij} = 1$. Intuitively, w_{ij} measures the intensity of neighbourhood between cities i and j. By convention, $w_{ii} = 0$.[8]

As in Lesage and Pace (2008), we can define the $N \times N$ row-standardized, destination-based spatial weight matrix $\mathbf{W}_d$, as $\mathbf{W}_d = \mathbf{I}_n \otimes \mathbf{W}$ or,

$$\mathbf{W}_d = \begin{pmatrix} \mathbf{W} & \mathbf{0}_n & \dots & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{W} & \dots & \vdots \\ \vdots & & \ddots & \mathbf{0}_n \\ \mathbf{0}_n & \dots & \mathbf{0}_n & \mathbf{W} \end{pmatrix}, \tag{3}$$

with $\mathbf{I}_n$ the $n \times n$ identity matrix and $\mathbf{O}_n$ an $n \times n$ matrix of zeros. This way, the spatial lag $N \times 1$ vector $\mathbf{W}_d\mathbf{y}$ contains the spatial average of air passenger flows from all neighboring destinations to each origin. It then introduces destination-based spatial dependence in the gravity model.

Similarly, we introduce the origin-based spatial dependence by forming the $N \times N$ row-standardized, origin-based spatial weight matrix $\mathbf{W_o}$, as $\mathbf{W_o} = \mathbf{W} \otimes \mathbf{I}_n$. The spatial lag of the dependent variable $\mathbf{W_o y}$ measures the connectivity relationship between all neighboring origins and each destination.

the origin city 2, n times to form the next n rows of $\mathbf{X}_o$ and so on.

[7] There is no consensus about how to best define the neighbourhood weight matrix and several alternative forms have been used in the literature. Overall, they depend in some way on the distance between the origin and destination. See Dubin (2003) for a discussion.

[8] Section 4 describes how we define the m-nearest neighbours.

Adding the spatial weight matrices $\mathbf{W_o}$ and $\mathbf{W}_d$ to (2), we define the following family of spatial autocorrelation models, which allow for spatial dependence, both in the air passenger flow vector $\mathbf{y}$ and the disturbance u,

$$\begin{aligned} \mathbf{y} &= \rho_o \mathbf{W_o y} + \rho_d \mathbf{W}_d \mathbf{y} + \alpha i_N + \beta_d X_d + \beta_o X_o + \gamma g + u \\ u &= \lambda_o \mathbf{W_o} u + \lambda_d \mathbf{W}_d u + \varepsilon, \end{aligned} \tag{4}$$

with $\varepsilon \sim N(0, \sigma_\varepsilon^2 \mathbf{I}_N)$. As discussed in Anselin (1988), the members of the family of spatial autocorrelation models can be derived from formulation (4). Setting $\lambda_o = \lambda_d = 0$ results in a LAG or "lagged autoregressive model", where the spatial dependence is modeled as occurring in the air passenger flow vector $\mathbf{y}$. $\mathbf{W_o y}$ then captures the origin-based spatial dependence, while $\mathbf{W}_d \mathbf{y}$ reflects the destination-based spatial dependence.[9]

In turn, the case where $\rho_o = \rho_d = 0$ yields a SEM or "Spatial Error model", where the disturbances follow a spatial autoregressive process.[10] Finally, a model where all ρ_o, ρ_d, λ_o and λ_d parameters are non zero implies a SAC or "Spatial Autocorrelation model", which allows for spatial autoregressive dependence both in air transport flows and the disturbances.[11]

By taking different assumptions on the strength of dependence parameters ρ_o, ρ_d, λ_o and λ_d and for simplicity's sake, we study seven special models of (4), as follows.

- **Model 1**: Assumption $\rho_o = \rho_d = \lambda_o = \lambda_d = 0$ yields the gravity model with independent observations of section 2.2.
- **Model 2**: Assumption $\rho_d = \lambda_o = \lambda_d = 0$ implies spatial dependence in the air passenger flow vector $\mathbf{y}$ and a single weight matrix $\mathbf{W_o}$, reflecting origin-based spatial dependence.
- **Model 3**: Assumption $\rho_o = \lambda_o = \lambda_d = 0$ results in another LAG model, with a single weight matrix $\mathbf{W}_d$, which captures autoregressive spatial dependence at destination.

[9] Focusing on this type of spatial dependence, Lesage and Pace (2008) consider a more general model:

$$\mathbf{y} = \rho_o \mathbf{W}_o + \rho_d \mathbf{W}_d + \rho_w \mathbf{W}_w + \alpha i_N + \beta_d \mathbf{X}_d + \beta_o \mathbf{X}_o + \gamma \mathbf{g} + \varepsilon,$$

with $\mathbf{W}_w = \mathbf{W}_o \cdot \mathbf{W}_d = \mathbf{I}_N - \rho_o \mathbf{W}_o - \rho_d \mathbf{W}_d + \rho_d \rho_o \mathbf{W}_d \cdot \mathbf{W}_o$.

The spatial weight matrix $\mathbf{W}_w$ reflects an average of flows from neighbours to the origin to neighbours to the destination.

[10] Focusing on spatial autocorrelation models which contain spatial errors, Dubin (2003)'s model writes as follows,

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\beta + u, \\ u &= \lambda \mathbf{W} u + \varepsilon, \end{aligned}$$

with $\mathbf{W}$ a spatial weight matrix.

[11] However, as stated by Dubin (2004), spatial autocorrelation models, with both a spatial lag and spatial error, are seldom used in practice, since it is very difficult to estimate them. See her Footnote 1.

- **Model 4**: Assumption $\rho_o = \rho_d$ and $\lambda_o = \lambda_d = 0$ also results in a LAG model, with a different single weight matrix, which we denote $\mathbf{W}_g$, with $\mathbf{W}_g \equiv \frac{1}{2}(\mathbf{W_o} + \mathbf{W}_d)$, reflecting a cumulative, non separable origin and destination spatial dependence effect.

- **Model 5**: Assumption $\rho_o = \rho_d = \lambda_d = 0$ allows for spatial autocorrelation in the errors, with a single weight matrix $\mathbf{W_o}$. λ_o then measures the intensity of the origin-based spatial autocorrelation of the disturbances.

- **Model 6**: Assumption $\rho_o = \rho_d = \lambda_o = 0$ differs from model 5 in the weight matrix $\mathbf{W}_d$ and the fact that λ_d measures the intensity of the destination-based spatial autocorrelation of the errors.

- **Model 7**: Assumption $\lambda_o = \lambda_d$ and $\rho_o = \rho_d = 0$ results in the weight matrix $\mathbf{W}_g$; the spatially auto correlated disturbances imply a cumulative, non separable origin and destination spatial dependence.

We rely on maximum-likelihood estimation procedures for the previous models, based on the technical results shown in Lesage and Pace (2008).

3 The data

Section 3 presents the dataset we use to apply spatial interaction models to air transport.

First, to measure OD air passenger flows, we rely on Sabre Airline Solutions' proprietary data intelligence solution, Global Demand Data (GDD), which provides air travel itineraries between airports all over the world, since 2002.[12] Specifically, we consider annual, OD, city to city air passengers, over the period $2010-2012$. The resulting dataset contains both economy and business passengers.[13]

Second, we use four explanatory variables, two of which are only available for a subset of 279 cities. The four explanatory variables are annual average air fares, gross domestic product (GDP), population per city and great-circle (GC) distance, per city-pair.

Sabre GDD provides information on air fares. Global Metromonitor 2012 provides information on GDP, population, employment and GDP per capita for 279 large metropolitan economies in the world, as measured by the size of their economies in 2010.[14] We consider real GDP at purchasing power parity (PPP) and population for these metro areas, over the period $2010-2012$.

[12] GDD aggregates information from world distribution systems, like Sabre, Amadeus and Galileo and performs adjustments to estimate total demand.

[13] We aggregate air passengers by city. Thus, we do not distinguish between cities with multiple airports.

[14] Global Metromonitor 2012 provides this information for 300 metro-areas. However, 21 of them have not been considered, due to one of the following reasons: lack of air traffic or because they were areas between two cities. See http://www.brookings.edu/research/reports/2012/11/30-global-metro-monitor for details.

Importantly, the availability of data on GDP and population constrains the number of OD air passenger flows to consider to $279^2 = 77841$ city-pairs.

Figure 1 and Table 1 show the representativeness per region of the 279 cities considered here.

Table 1: The 279 cities considered

Region	Number of cities	Icon in the map
Asia/Pacific	85	△
Europe	78	×
North America	81	◯
CIS	3	•
Africa	7	◇
Middle East	5	▲
Latin America	20	■

Figure 1: Geographical representativeness of the 279 cities

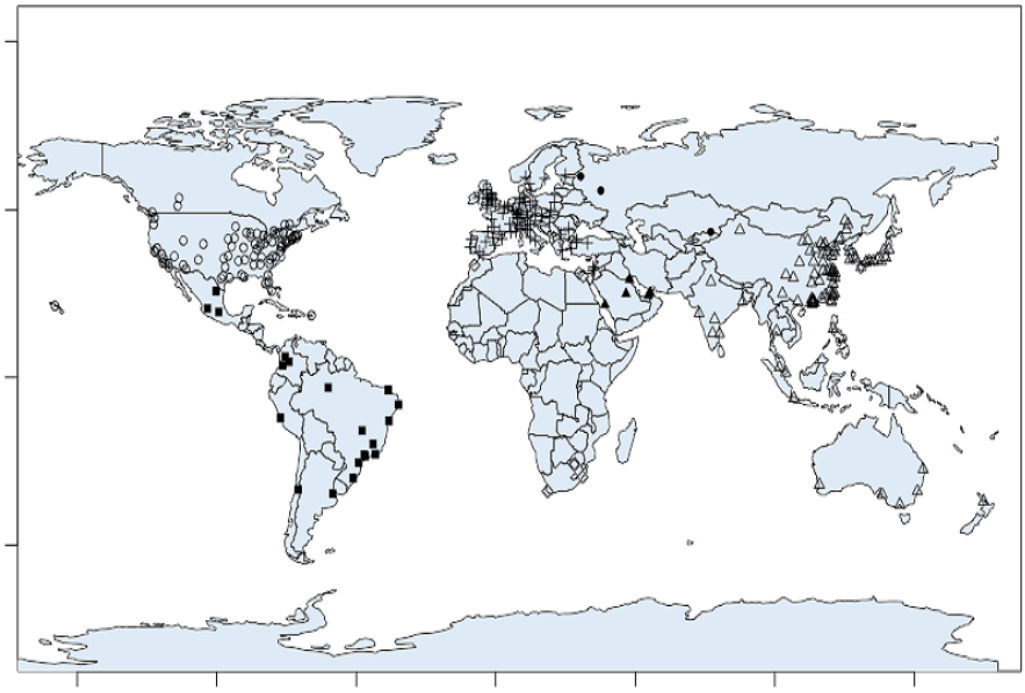


Table 2 presents the descriptive statistics of the variables under study, in 2012.

Table 2: Descriptive statistics, 2012

	Air passengers	Population (thousands)	Nominal GDP (mill $, PPP)	GC distance (km)	Annual average air fares ($)
Mean	15526	4565	135192	7698	754
Median	106	2745	83639	8173	620
Std Dev	89714	5248	164134	4444	886
N	77841	77841	77841	77841	77841

4 Model estimations

Relying on the data presented in section 3, section 4 estimates the family of seven models introduced in section 2.3. Section 4 has three parts. The first part estimates model 1, assuming independent observations, that is, $\rho_o = \rho_d = \lambda_o = \lambda_d = 0$ in equation (4). The second part tests the absence of spatial dependence in air passenger flows. Finally, by creating the two spatial connectivity matrices for origin and destination spatial dependence, $\mathbf{W}_o$ and $\mathbf{W}_d$ respectively, the third part estimates the remaining six models.

The air passenger flow matrix (1), one for each time period, contains annual air passengers from each of the n cities of origin to each of the n destination cities, over the period $2010-2012$.[15] We then transform each air passenger flow matrix, using $\log(vec(\mathbf{Y})) = \log(\mathbf{y})$, to produce a cross-sectional vector, representing the logged air passenger flows.

4.1 Model 1, assuming independent observations

We consider four alternative specifications of model 1 ($\rho_o = \rho_d = \lambda_o = \lambda_d = 0$). In specification $(1.a)$, the two explanatory variables are real GDP, at PPP, and population. After eliminating all zero-flows and due to appropriate transformations of the matrix $\mathbf{X}$ of explanatory variables, we obtain the $N \times 2$ matrices $\mathbf{X}_o$ and $\mathbf{X}_d$, containing the GDP and the population of the origin and destination cities, respectively. Also, we create the $N \times 1$ vector $\log(\mathbf{g}) \equiv \log(vec(\mathbf{G}))$, containing the distances from each city of origin to each destination city.

Specification $(1.b)$ adds to $(1.a)$ 12 indicator variables, one for each origin and destination region. X_o and X_d then become $N \times 14$ matrices. In turn, specification $(1.c)$ adds (logged) average air fares to $(1.b)$. Finally, instead of eliminating the zero passenger flows as in $(1.a)$ to $(1.c)$, specification $(1.d)$ modifies the dependent variable, using $\log(1 + \mathbf{y})$, and then introduces an indicator variable, which takes the value of 1 if the passenger flow is 1.

Since the presence of zero air passenger flows is one of the typical problems that arise in applied practice, we briefly discuss it here. Several approaches exist to deal with this issue: Zero flow

[15] The number of n origin and destination cities varies from year to year.

elimination (provided the number of these flows is not too large); modification of the dependent variable, using $\log(1 + \mathbf{y})$ to accommodate the log transformation, and Poisson regressions.

As mentioned, we follow the first and second approach, that is, specifications $(1.a)$ to $(1.c)$ eliminate the zero air passenger flows, while specification $(1.d)$ modifies the dependent variable, using $\log(1 + \mathbf{y})$ and introduces the indicator variable for the zero counts. The reasons for this choice follow.

First, the number of zero counts, representing 22% of total flows in 2012,[16] does not invalidate the use of least-squares regressions. Second, introducing the indicator variable for zero flows (after the $\log(1+\mathbf{y})$ transformation) allows us to measure whether the non-availability of a flight between two cities (that is, the case of a zero flow) may affect the estimate results. Third, Poisson regression is mostly used when there is a large proportion of zero flows. By large, Fisher and Lesage (2010) mean greater than 50% to 70% of total flows. We are far from these percents.

Another difficulty that arises when estimating flow data is the treatment of intra-regional flows and inter-regional flows.[17] Since intra-regional flows tend to be considerably larger than inter-regional flows, two common practices exist to deal with them. First, set the flows in the main diagonal to zero[18]. Second, as Lesage and Pace (2008) propose, create separate models for each type of flow. The latter is to avoid that large intra-regional flows excessively influence the coefficient estimates of the origin and destination explanatory variables.

In contrast, we do not need to choose between these procedures, because, by definition, flows in the main diagonal of the air passenger flow matrix (1), representing air passengers within cities, are zero.

Table 3 presents the results of the four specifications of model 1, for 2012. The model estimations for 2010 and 2011 are in the appendix.

[16] The proportion of zero counts represents 23% of total flows, both in 2010 and 2011.

[17] Intra-regional flows are flows within the region, which are in the main diagonal of the flow matrix; inter-regional flows are flows between regions.

[18] See Tiefelsdorf (2003) and Fischer and Lesage (2010).

Table 3: Model estimations with independent observations, 2012

Variable	$(1.a)$	$(1.b)$	$(1.c)$	$(1.d)$
Constant	-10.386^{***}	-12.399^{***}	-11.845^{***}	-7.403^{***}
$\log(GDP_o)$	1.448^{***}	1.083^{***}	1.207^{***}	1.072^{***}
$\log(POP_o)$	-0.231^{***}	0.351^{***}	0.215^{***}	0.078^{***}
$\log(GDP_d)$	1.464^{***}	1.088^{***}	1.214^{***}	1.076^{***}
$\log(POP_d)$	-0.256^{***}	0.337^{***}	0.197^{***}	0.061^{***}
$\log(vec(G))$	-1.530^{***}	-1.465^{***}	-0.756^{***}	-0.283^{***}
$I^O_{Asia\ Pacific}$		-0.601^{***}	-0.737^{***}	-0.656^{***}
I^O_{Europe}		0.562^{***}	0.203^{***}	-0.005
$I^O_{NorthAmerica}$		0.823^{***}	0.514^{***}	0.258^{***}
I^O_{CIS}		0.009	-0.294^{**}	-0.400^{***}
I^O_{Africa}		0.853^{***}	0.701^{***}	0.432^{***}
$I^O_{Middle\ East}$		0.395^{***}	0.412^{***}	0.283^{***}
$I^D_{Asia/Pacific}$		-0.630^{***}	-0.776^{***}	-0.690^{***}
I^D_{Europe}		0.511^{***}	0.163^{***}	-0.029
$I^D_{North\ America}$		0.827^{***}	0.524^{***}	0.263^{***}
I^D_{CIS}		-0.045	-0.343^{***}	-0.446^{***}
I^D_{Africa}		0.808^{***}	0.663^{***}	0.403^{***}
$I^D_{Middle\ East}$		0.358^{***}	0.375^{***}	0.249^{***}
$\log(Fare)$			-1.039^{***}	-1.477^{***}
1{Zero Flow}				-14.889^{***}
Observations	60188	60188	60188	77841
Adjusted R^2	0.411	0.454	0.475	0.696
AIC	277597	273071	270641	332488
P-value F-Stat	< 0.001	< 0.001	< 0.001	< 0.001

Note. Level of significancy : * 10% , ** 5 % , *** 1%.

With four exceptions, the variable estimates in table 3 are significant at the usual confidence levels. Distance and air fares are always significant and negative, whereas GDP and population, both at the origin and destination, are positive, with two exceptions in specification $(1.a)$. Based on the adjusted R^2, we opt for specification $(1.d)$.[19]

A characteristic of the previously estimated models is that changes in the value of an explanatory variable associated with a city will potentially impact air passenger flows to other cities. As an

[19] Fisher and Lesage (2010) note that using $\log(1+y)$ may potentially lead to downward bias in the coefficient estimates. Nevertheless, we do not find any evidence of this downward bias. Also, we prefer specification $(1.d)$, because it allows us to compare easily the traditional gravity model with models that account for spatial dependence (section 4.3).

example, a *ceteris paribus* 1% decrease in the explanatory variable GDP in city i implies that city i will be viewed differently, both as an origin and a destination. Given matrices $\mathbf{X}_d$ and $\mathbf{X}_o$, the -1% of GDP of city i will result in changes of $2n$ observations of the explanatory variable matrices.

Continuing with the example, an estimated 1.072 coefficient for GDP at origin in specification $(1.d)$ means that due to the 1% downside in city i economy, residents of city i will be less propense to travel by air, because of the wealth effect. The city will then exert less *push*, leading to an expected 1.072% decrease in air traffic from this city. Also, city i will exert less *pull,* resulting in a predicted 1.076% drop in air traffic to this city.

Interestingly, indicator variables for regions, at the origin and destination, are significant, with 2 exceptions. This suggests that overall, spatial heterogeneity across regions is relevant.[20] Finally, the estimated coefficient for the indicator variable $1\{\text{Zero Flow}\}$ downscales the effect that zero flows have on OD air passengers.

4.2 Testing the absence of spatial autocorrelation

We start by specifying the neighbourhood weight matrix, $\mathbf{W}$. We then test the absence of spatial autocorrelation for the OD, city pair, air passenger flows and the residuals of model $(1.d)$ estimation, using a Moran test.

To compute $\mathbf{W}$, we rely on the method of the m nearest neighbours, in terms of great circle distance, with $m = 3$.[21]

4.2.1 Moran's test for the OD air passenger flow

Since the distribution of the OD air passenger flows is non Gaussian, we use the non-free sampling (randomization) version of the Moran Test, with 1000 permutations.[22]

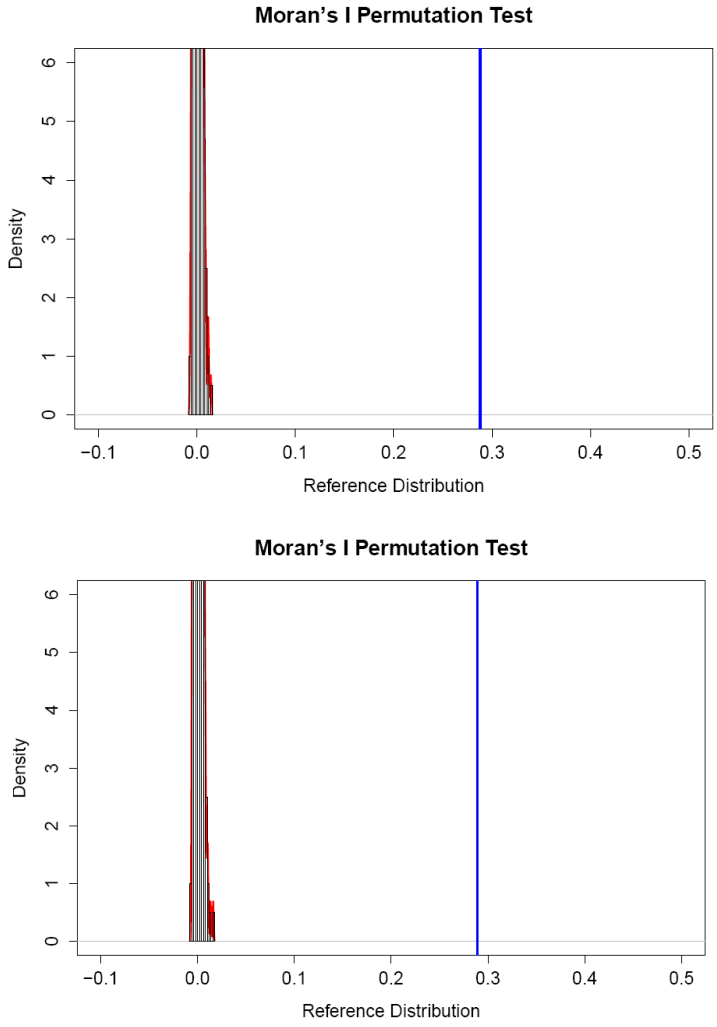
Figure 2 shows the density plots of the permutation outcomes, for the neighbour matrixes $\mathbf{W}_o$ (top) and $\mathbf{W}_d$ (bottom).

[20] Spatial heterogeneity refers to the variation of OD air passengers across regions. It implies that parameters vary by location.

[21] After trying other values for m, we choose $m = 3$, since it avoids having abnormal neighbours, from a geographic point of view. See the appendix for the geographic representation of the neighbourhood matrix.

[22] For a fixed neighbourhood matrix W_d or W_o, the non-free sampling version of the Moran Test consists of randomly drawing T permutations (here $T = 1000$) of the cross sectional air passenger vector y, computing the Moran index I for each permutation and the $I_{\min}, I_{\max}$. It then compares the observed Moran's I with the interval $[I_{\min}, I_{\max}]$. We reject H_o if Moran's I does not belong to this interval.

Figure 2: Moran test for air passenger flows (non-free sampling version)



Note. Density plots of the permutation outcomes, for the neighbour matrixes W_o (top) and W_d (bottom).

Because both Moran indexes are far above their intervals $[I_{\min}, I_{\max}]$, we reject the null of absence of spatial autocorrelation.

4.2.2 Moran's test for the residuals of the gravity model ($1.d$)

Table 4 shows the results of the free-sampling version of the Moran test for the residuals of specification ($1.d$).

Table 4: Moran test for the residuals of the gravity model ($1.d$) (free-sampling version)

Variable	Moran's I statistic	$P-value$
Test with $\mathbf{W}_o$	144.39	< 0.001
Test with $\mathbf{W}_d$	146.53	< 0.001

As before, we reject the null of absence of spatial autocorrelation. We conclude that least square estimates and inferences that ignore the spatial dependence present in our data are not justified. In the next section, we estimate spatial interaction models that allow for spatial dependence (models 2 to 7, as defined in section 2.3).

4.3 Models 2 to 7, with spatial dependence

Taking different assumptions on the strength of the dependence parameters, ρ_o, ρ_d, λ_o and λ_d table 5 presents the estimation results. As stated in section 2.3, models 2 to 4 allow for spatial dependence in the air passenger flow vector $\mathbf{y}$, while models 5 to 7 allow for spatial autocorrelation in the disturbances. All models have a single weight matrix.

Table 5: Model estimations with spatial dependence, 2012

Variable	*Model* 2	*Model* 3	*Model* 4	*Model* 5	*Model* 6	*Model* 7
Constant	-10.748^{***}	-10.706^{***}	-13.466^{***}	-13.961^{***}	-13.983^{***}	-27.258^{***}
$\log(GDP_o)$	1.324^{***}	0.714^{***}	0.978^{***}	1.381^{***}	1.129^{***}	1.541^{***}
$\log(POP_o)$	0.042^{***}	-0.051^{***}	-0.072^{***}	-0.008	0.175^{***}	0.020
$\log(GDP_d)$	0.723^{***}	1.324^{***}	0.979^{***}	1.125^{***}	1.377^{***}	1.527^{***}
$\log(POP_d)$	-0.066^{***}	0.031^{**}	-0.081^{***}	0.165^{***}	-0.015	0.022^{***}
$\log(vec(G))$	-0.050^{***}	-0.046^{***}	0.146^{***}	-0.198^{***}	-0.192^{***}	0.334^{***}
$I^O_{Asia\ Pacific}$	-0.401^{***}	-0.332^{***}	-0.128^{***}	-0.598^{***}	-0.700^{***}	-0.158^{***}
I^O_{Europe}	0.270^{***}	-0.112^{***}	0.149^{***}	0.430^{***}	0.290^{***}	1.699^{***}
$I^O_{NorthAmerica}$	0.166^{***}	0.013	-0.049^{*}	0.470^{***}	0.493^{***}	1.199^{***}
I^O_{CIS}	-0.636^{***}	-0.396^{***}	-0.613^{***}	-0.395^{***}	-0.141	0.284^{***}
I^O_{Africa}	0.698^{***}	0.324^{***}	0.577^{***}	0.290^{***}	0.454^{***}	1.212^{***}
$I^O_{Middle\ East}$	-0.097^{*}	0.114^{**}	-0.218^{***}	0.262^{***}	0.426^{***}	0.620^{***}
$I^D_{Asia/Pacific}$	-0.373^{***}	-0.438^{***}	-0.170^{***}	-0.719^{***}	-0.613^{***}	-0.164
I^D_{Europe}	-0.132^{***}	0.247^{***}	0.127^{***}	0.272^{***}	0.421^{***}	1.682^{***}
$I^D_{North\ America}$	0.011	0.163^{***}	-0.587^{**}	0.509^{***}	0.492^{***}	1.228^{***}
I^D_{CIS}	-0.427^{***}	-0.684^{***}	-0.646^{***}	-0.184^{***}	-0.430^{***}	0.247
I^D_{Africa}	0.303^{***}	0.673^{***}	0.556^{***}	0.420^{***}	0.259^{***}	1.173^{***}
$I^D_{Middle\ East}$	0.087	-0.122^{**}	-0.236^{***}	0.407^{***}	0.236^{***}	0.614^{***}
$\log(Fare)$	-1.195^{***}	-1.199^{***}	-0.966^{***}	-1.301^{***}	-1.308^{***}	-0.947^{***}
1{Flow Nul}	-12.145^{***}	-12.156^{***}	-9.893^{***}	-12.865^{***}	-12.910^{***}	-9.763^{***}
ρ_o	0.358^{***}					
ρ_d		0.360^{***}				
ρ_w			0.655^{***}			
λ_o				0.516^{***}		
λ_d					0.520^{***}	
λ_w						0.877^{***}
Adjusted R^2	0.755	0.755	0.804	0.760	0.761	0.812
AIC	315850	315561	298340	314000	313520	295030
P-value F-Stat	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Notes. Level of significancy : * 10% , ** 5 % , *** 1%.

Observations, models 2 to 7: 77841.

Table 5 shows that population, GDP (both at the origin and destination), annual average air fares and great-circle distance continue to be significant, at usual confidence levels. However, a direct comparison of the values of the coefficients from the least-square estimates and the spatial

models 2 to 4 is not valid (see Lesage and Pace (2008) and Lesage and Thomas-Agnan (2014))[23]

If we first focus on models 2 to 4, the estimates for $\rho_o = 0.358$ and $\rho_d = 0.360$ indicate spatial dependence of almost equal importance, between neighbours to the origin and neighbours to the destination. It then provides evidence in favour of a cumulative, non separable origin and destination spatial dependence effect, as captured in model 4.

We obtain the same conclusion if instead, we focus on models 5 to 7, that is, there is evidence of a cumulative, non separable origin and destination spatial dependence effect in the disturbances. Importantly, table 5 shows high levels of spatial dependence, regardless of the spatial model considered.

We now compare the spatial models in table 5 with the gravity model, specification $(1.d)$. Table 6 displays the likelihood ratio (LR) tests of model $(1.d)$, versus each spatial model, models 2 to 7.[24] Also, it displays the Akaike criteria for the seven model estimates.

Table 6: Log likelihood of the gravity model $(1.d)$ and spatial models (models 2 to 7)

Model	LR test versus model $(1.d)$	Critical value $\alpha = 0.05$	AIC
$1.d$			332488
2	16642	3.84	315850
3	16939	3.84	315561
4	34142	3.84	298340
5	18490	3.84	314000
6	18968	3.84	313520
7	37457	3.84	295030

As reflected in table 6, when comparing any of the models that account for spatial dependence (models 2 to 7) with the least-square estimate (model $1.d$), we conclude that any spatial model is better than model $(1.d)$, since they all have smaller AIC than model $(1.d)$. Also, each LR test rejects model $(1.d)$, in favor of the spatial model. This is still true if instead, we conduct a Lagrange Multiplier test.

Moreover, model 7, which reflects a cumulative, non-separable, origin and destination spatial dependence effect in the disturbances, seems to dominate all other models, as it has the smallest AIC. However, a definite statement regarding the choice of the model exceeds the scope of this paper. The reasons follow.

[23] This is because evaluating the impact of explanatory variables in spatial autoregressive model requires the methodology introduced in Lesage and Thomas-Agnan (2014).

[24] For the LR tests, we calculate the statistic $2(L_U - L_R)$, where L_U is the likelihood function of the spatial model and L_R the likelihood function of model $(1.d)$. This statistic is asymptotically distributed as a $\varkappa^2$ random variable, with degrees of freedom equal to the number of restricted parameters.

As discussed in Dubin (2003), there is little theoretical justification for the choice between spatial models and very often researchers choose the model that predicts the best. Since in this paper we focus on the estimation of the parameters of the OD air passenger model, we can not choose between models, based on their predictive ability.

More generally, the aim of this paper has been to assess whether spatial autocorrelation matters, when estimating OD air passengers. We find significant evidence of spatial dependence in air passenger flows, both at origin, at destination and at origin and destination. Thus, contrary to common practice, least-square estimates and inferences that ignore spatial dependence seem not to be justified.

5 Practice and Policy

This paper takes a step towards improving our understanding of modeling OD air passengers, by explicitly taking into account spatial autocorrelation. This is crucial, because being able to estimate the number of air passengers between two cities at a given point in time is of major importance, both for aircraft manufacturers and airlines.

Aircraft manufacturers, such as Airbus, rely on this type of modelling to assess the future demand for civil passenger and freighter aircraft, which in turn, steer them towards innovation. Airlines also need these forecasts to decide whether to open new routes, offer more frequencies and/or increase aircraft capacity.

As an illustration, this section describes how this type of modeling helps Airbus to assess the future demand for civil passenger and freighter aircraft, in the context of the Global Market Forecast (GMF) methodology.

The GMF consists of three main steps: the traffic forecast giving the overall shape of the expected traffic evolution over the next 20 years; the network forecast, identifying the future evolution of the airlines' networks and finally, the demand forecast, estimating the number of aircraft which will be required to accommodate the expected traffic growth.

Modeling OD air passengers is part of the second previously stated step. More specifically, the network forecast step relies on an in-house network-planning model[25] to determine how many passengers will fly over the next 20 years, which itineraries they are likely to choose and when and where airlines will respond to the expected passenger evolution, by opening or removing routes.

Airbus' network-planning model starts by breaking down the traffic forecast between country pairs, down to the estimation of OD, city-pair air passengers. Second, for each airline, it constructs the itineraries (routes), that is, a flight or sequence of flights used to travel between any two cities.[26]

[25] A network-planning model is a collection of sub-models, to be described. See Garrow (2010) for a detailed description.

[26] As it is typical in this type of applications, itineraries are limited to non-stop, single and double connections.

Importantly, itineraries do not only include existing routes, but also future route candidates.[27]

Third, a market share model allows predicting the percentage of travellers that are likely to select each itinerary, existing or new, at each city-pair and at each point in time (year). The market share model Airbus uses is a "Quality of service index" (QSI) model.[28] In order to determine the share of each itinerary on the OD city-pair, the QSI model considers attributes like flight frequency, type of connection and circuitry, as quality of services.

Finally, the demand of each itinerary is determined by multiplying the percentage of travellers expected to travel on each itinerary by the expected market size, that is, the number of OD, city-pair air passengers. The importance of adequately estimating the number of air passengers between any two cities becomes now clear, as it gives the size of the OD city pair. This, in turn, enables Airbus to predict when and where airlines are likely to open or remove a route, and this way, how the shape of the airlines network is likely to evolve through time.

6 Conclusion

In this paper, we take a step towards improving our understanding of modeling origin-destination, city-pair, air passengers, by explicitly taking into account spatial autocorrelation. One empirical question motivates us, that is, whether the characteristics at proximal cities impact air passenger flows, between two cities.

The literature has extensively used gravity models to estimate air passenger flows. However, the main particularity of these models is that they assume spatial independence between origin-destination pairs. More specifically, they suppose that the distance between the origin and the destination can effectively eliminate the spatial structure, potentially present in origin-destination flow data.

To challenge this assumption, we build on Dubin (2003) and Lesage and Pace (2008) and modify the traditional gravity model, to account for spatial dependence, both in air passenger flows and the disturbances.

We estimate six spatial models, which allow for spatial autoregressive dependence (spatial lag) or spatially auto correlated error term (spatial error). Based on likelihood ratio tests and informational criteria, we conclude that any of the spatial models considered here is better than the traditional gravity model. This implies that, contrary to common practice, least-square estimates and inferences that ignore spatial dependence seem not to be justified.

[27] The identification of new route candidates considers airlines' current network and the potential size of new markets.

[28] As defined by Garrow (2010), QSI models relate an itinerary's passenger share to its "quality" (and the quality of all other itineraries in the city or airport pair). Quality is defined as a function of various itinerary service attributes and their corresponding preference weights. See Garrow (2010) for details.

Interestingly, we find that the model which reflects a cumulative, origin and destination spatial dependence effect in the disturbances seems to be the most appropriate, based on the same aforementioned criteria. It is important to stress though, that we reach this conclusion, by focusing on the explanatory aspects and not on the predictive ones.

If instead, the focus were on prediction, we would need appropriate prediction formulae for spatial flow models (see Goulard, *et.al* (2013) for the case of prediction on spatial autoregressive models). This constitutes a future venue of research, that is, to compare the spatial model estimates considered here, based on their predictive ability. It will be the topic of forthcoming research.

References

[1] Bhadra D, Kee J (2008) Structure and dynamics of the core US air travel markets: A basic empirical analysis of domestic passenger demand. Journal of Air Transport Management 14 (1): 27-39.

[2] Boros A, Lee J, Shaw S L (1993) Analysis of network structure for US domestic airline networks. Papers and Proceedings of Applied Geography Conference 16: 136-143.

[3] Curry L (1972) Spatial analysis of gravity flows. Regional Studies 6: 131-147.

[4] Doganis R (2004) Flying Off Course: The economics of international airlines. Third ed. Routledge, London, New York.

[5] Dubin R (2003) Robustness of spatial autocorrelation specifications: some Monte Carlo evidence. Journal of Regional Science 43 (2): 221-248.

[6] Dubin R (2004) Spatial lags and spatial errors revisited: Some Monte Carlo evidence. Advances in Econometrics 18: 75-98.

[7] Fischer M, LeSage J (2010) Spatial econometric methods for modeling origin-destination flows. Handbook of applied spatial analysis: Software tools, methods and applications, eds Fischer M, Getis A: 409-433. Springer-Verlag Berlin Heidelberg.

[8] Fotheringham A S (1981) Spatial structure and distance-decay parameters. Annals of the Association of American Geographers 71 (3): 425-436.

[9] Garrow L (2010) Discrete choice modelling and air travel demand: theory and applications. Ashgate Publishing Limited, England, and Ashgate Publishing Company, USA.

[10] Goulard M, Laurent T, Thomas-Agnan C (2013) About predictions in spatial autoregressive models: Optimal and almost optimal strategies. Toulouse School of Economics Working Paper, 13 (452), pp??.

[11] Griffith, D. A. and Jones, K.G. (1980) Explorations into the relationship between spatial structure and spatial interaction. Environment and Planning A, 12 (2), 187-201.

[12] Griffith, D. A. (2007) Spatial structure and spatial interaction: 25 Years Later. The Review of Regional Studies, 37(1), 28-38.

[13] Grosche, T., Rothlauf, F. and Heinzl, A. (2007) Gravity models for airline passenger volume estimation. Journal of Air Transport Management, 13, 175-183.

[14] Jorge-Calderon, J.D. (1997) A demand model for scheduled airline services on international European routes. Journal of Air Transport Management, 3 (1), 23-35.

[15] Kockelman, K.M., Wang X. and Wang Y. (2013) Understanding spatial filtering for analysis of land use-transport data. Journal of Transport Geography, 31, 123-131.

[16] LeSage, J.P. and Pace, R. K. (2004) Spatial and spatiotemporal econometrics. Advances in econometrics, 18, 1-32.

[17] LeSage, J.P. and Pace, R. K. (2008) Spatial econometric modeling of origin-destination flows. Journal of Regional Science, 48 (5), 941-67.

[18] LeSage J. P. and Pace, R. K. (2010) Spatial econometric models. Handbook of applied spatial analysis: Software tools, methods and applications, eds Fischer, M. and Getis, A., 355-376. Springer-Verlag Berlin Heidelberg.

[19] LeSage J.P. and Thomas-Agnan, C. (2014) Spatial econometric origin-destination flow models. Forthcoming in Handbook of Regional Science, 1653-1673.

[20] Russon, M. and Riley, N. (1993) Airport substitution in a short haul model of air transportation. International Journal of Transportation Economics, 20, 157-173.

[21] Tiefelsdorf, M. (2003) Misspecifications in interaction model distance decay relations: A spatial structure effect. Journal of Geographical Systems, 5, 25–50.

7 Appendix

Table 7: Model estimations with independent observations, 2010

Variable	$(1.a)$	$(1.b)$	$(1.c)$	$(1.d)$
Constant	-11.832^{***}	-13.168^{***}	-12.279^{***}	-7.147^{***}
$\log(GDP_o)$	1.540^{***}	1.192^{***}	1.385^{***}	1.178^{***}
$\log(POP_o)$	-0.293^{***}	0.236^{***}	0.026	-0.036^{*}
$\log(GDP_d)$	1.535^{***}	1.169^{***}	1.361^{***}	1.156^{***}
$\log(POP_d)$	-0.283^{***}	0.267^{**}	0.055^{*}	-0.013
$\log(vec(G))$	-1.488^{***}	-1.435^{***}	-0.415^{***}	-0.192^{***}
$I^O_{Asia\ Pacific}$		-0.634^{***}	-0.639^{***}	-0.525^{***}
I^O_{Europe}		0.045^{***}	0.052	-0.041
$I^O_{NorthAmerica}$		0.048^{***}	0.281^{***}	0.130^{***}
I^O_{CIS}		-0.146	-0.309^{***}	-0.307^{***}
I^O_{Africa}		0.810^{***}	0.800^{***}	0.547^{***}
$I^O_{Middle\ East}$		0.228^{**}	0.459^{***}	0.332^{***}
$I^D_{Asia/Pacific}$		-0.604^{***}	-0.623^{***}	-0.515^{***}
I^D_{Europe}		0.420^{***}	0.103^{*}	0.001
$I^D_{North\ America}$		0.766^{***}	0.375^{***}	0.203^{***}
I^D_{CIS}		-0.123	-0.300^{***}	-0.310^{***}
I^D_{Africa}		0.885^{***}	0.891^{***}	0.611^{***}
$I^D_{Middle\ East}$		0.255^{**}	0.463^{***}	0.330^{***}
$\log(Fare)$			-1.613^{***}	-1.791^{***}
1{Flow Nul}				-16.47^{***}
Observations	59738	59738	59738	77841
Adjusted R^2	0.4152	0.4506	0.5018	0.7128
AIC	275340.5	271615.7	265770.5	327343.2
P-value F-Stat	< 0.001	< 0.001	< 0.001	< 0.001

Note. Level of significancy : * 10% , ** 5 % , *** 1%.

Table 8: Model estimations with independent observations, 2011

Variable	$(1.a)$	$(1.b)$	$(1.c)$	$(1.d)$
Constant	-11.464^{***}	-12.798^{***}	-11.909^{***}	-7.171^{***}
$\log(GDP_o)$	1.502^{***}	1.131^{***}	1.291^{***}	1.130^{***}
$\log(POP_o)$	-0.256^{***}	0.299^{***}	0.124^{***}	0.012
$\log(GDP_d)$	1.514^{***}	1.122^{***}	1.277^{***}	1.121^{***}
$\log(POP_d)$	-0.274^{***}	0.306^{**}	0.133^{***}	0.015
$\log(vec(G))$	-1.489^{***}	-1.436^{***}	-0.632^{***}	-0.248^{***}
$I^O_{Asia\ Pacific}$		-0.576^{***}	-0.755^{***}	-0.667^{***}
I^O_{Europe}		0.467^{***}	0.074	-0.087^{*}
$I^O_{NorthAmerica}$		0.786^{***}	0.379^{***}	0.141^{***}
I^O_{CIS}		-0.026	-0.364^{***}	-0.435^{***}
I^O_{Africa}		0.836^{***}	0.676^{***}	0.396^{***}
$I^O_{Middle\ East}$		0.367^{***}	0.372^{***}	0.234^{***}
$I^D_{Asia/Pacific}$		-0.590^{***}	-0.783^{***}	-0.684^{***}
I^D_{Europe}		0.473^{***}	0.090^{*}	-0.068^{*}
$I^D_{North\ America}$		0.823^{***}	0.416^{***}	0.169^{***}
I^D_{CIS}		-0.038	-0.365^{***}	-0.426^{***}
I^D_{Africa}		0.817^{***}	0.672^{***}	0.408^{***}
$I^D_{Middle\ East}$		0.347^{***}	0.349^{***}	0.216^{***}
$\log(Fare)$			-1.232^{***}	-1.583^{***}
$1\{\text{Flow Nul}\}$				-15.459^{***}
Observations	59899	59899	59899	77841
Adjusted R^2	0.4138	0.451	0.4821	0.7026
AIC	275909	272024.4	268493	330690.9
P-value F-Stat	< 0.001	< 0.001	< 0.001	< 0.001

Note. Level of significancy : * 10% , ** 5 % , *** 1%.

Figure 4: Histogram of residuals of model (1.*c*) estimation, 2012

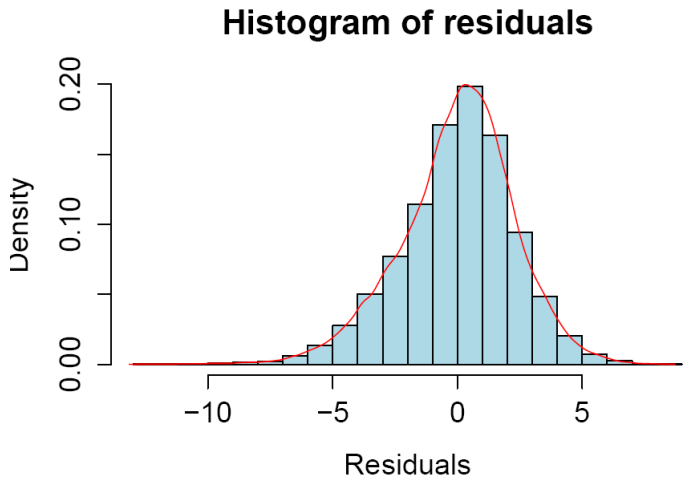


Figure 5: Geographic representation of the neighbourhood matrix, W